AD-A070 012 WISCONSIN UNIV-MADISON DEPT OF STATISTICS F/G 12/1

DESIGN PROBLEMS FOR OPTIMAL SURFACE INTERPOLATION.(U)

MAY 79 C A MICCHELLI, G WAHBA N00014-77-C-0675

UNCLASSIFIED UWIS-DS-79-565 NL

1 OF 1

AD A070012

END

DATE FILMED

1-82

DTIC

Department of Statistics
University of Wisconsin

TECHNICAL REPORT NO. 565

May 1979

DESIGN PROBLEMS FOR OPTIMAL SURFACE INTERPOLATION

Charles A. Micchelli
IBM
T. J. Watson Research Center
Yorktown Heights, NY 10598

and

Grace Wahba
University of Wisconsin
Department of Statistics
1210 West Dayton Street
Madison, Wisconsin 53706

Design problems for optimal surface interpolation

Charles A. Micchelli

IBM
T. J. Watson Research Center
Yorktown Heights, NY 10598

Grace Wahba *

Department of Statistics
University of Wisconsin
Madison, Wisconsin

Typed by Martha Cooper
Formatted using the Yorktown Formatting Language
Printed in the Experimental Printer

ABSTRACT

We consider the problem of interpolating a surface given its values at a finite number of points. We place a special emphasis on the question of choosing the location of the points where the function will be sampled.

Using minimal norm interpolation in reproducing kernel Hilbert spaces, equivalently Bayesian interpolation, and N-widths, we provide lower bounds for interpolation error relative to certain error criteria. These lower bounds can be used when evaluating an existing design, or when attempting to obtain a good design by iterative procedures to decide whether further minimization is worthwhile. The bounds are given in terms of the eigenvalues of a relevant reproducing kernel and the asymptotic behavior of these eigenvalues for certain tensor product spaces in the unit d-dimensional cube is obtained.

We demonstrate that for H_m, the d-dimensional tensor product of Sobolev spaces $W_2^{(m)}[0,1]$ and $P_N g$, the minimal norm interpolant to g at N given data points, the uniform convergence of $||g - P_N g||_{H_m}$ over g in the unit ball in H_{2m} cannot proceed at a rate faster than $((\log N)^{d-1}/N)^{2m}$. Certain conjectures concerning designs converging at this rate are made.

*The research of this author was supported by the Office of Naval Research under Grant No. N00014-77-C-0675.

1. Introduction.

We are interested in the problem of recovering a surface $g(t), t \in T$, from observations of g at a discrete set $T_N = \{t_i\}_{i=1}^{N}$ of points in T (called the "design"). In particular, we are interested in choosing T_N so that an estimate, say, g_N of g from the data $\{g(t_i)\}_{i=1}^{N}$ is closest to g in some appropriate sense among all designs T_N.

This problem arises in numerous applications. To cite one group of examples, T may be a sphere (the surface of the earth) or a rectangle and g(t) the 500 millibar height or the temperature, or the concentration of some air pollutant at position t. The interpolation problem requires an estimate of g over the entire surface given its values on T_N while the design problem concerns optimal or nearly optimal choices of T_N.

In this introduction we shall briefly survey several different ways of viewing the interpolation problem i.e. reconstructing the function from its sample values, and then follow this discussion with a description of some known results for the design problem in one dimension.

In this discussion of interpolation we will distinguish between the Bayesian approach and the function-analytic or deterministic approach. We further distinguish the problem of estimating g(t), for all $t \in T$ from estimating g(t) at a single point in T as well as introduce the possibility that our observations are distorted with errors. However, this latter feature is not primary for our objectives here.

The Bayesian approach is as follows: We suppose g(t), $t \in T$, is a Gaussian stochastic process, or "random field" with zero mean and given strictly positive definite (prior) covariance $K(s,t) = Eg(s)g(t)$, $s,t \in T$. Given the data $g(t_1),\dots,g(t_N)$ the Bayesian estimator for g(t) is

$$g_N(t) = E\{g(t)\,|\,g(t_1),\dots,g(t_N)\}$$

$$= (K_{t_1}(t),\dots,K_{t_N}(t))K_N^{-1}\begin{pmatrix} g(t_1) \\ \cdot \\ \cdot \\ \cdot \\ g(t_N) \end{pmatrix}$$

where $K_{t_i}(t) = K(t_i,t)$ and K_N is the $N \times N$ matrix with i, jth entry $K(t_i,t_j)$, and thus

$$\min_a E(g(t)-\sum_{i=1}^{N} a_i g(t_i))^2$$

$$= E(g(t)-g_N(t))^2.$$

The functional analytic approach is closely related to the Bayesian approach . Instead of assuming that g is a stochastic process, suppose g is a fixed element of H_K, the reproducing kernel Hilbert space with space reproducing kernel K.

Then g_N may be shown to be the minimal norm interpolant to g on T_N in H_K, the Hilbert space with reproducing kernel K. Observing that $<g,K_{t_i}>_K = g(t_i)$ where $<\cdot,\cdot>_K$ is the inner product in H_K it can be verified that if $P_N g$ is the minimal norm interpolator of g on T_N i.e.

$$||P_N g||_K = \min_{h(t_i)=g(t_i)} ||h||_K, \qquad (P_N g)(t_i) = g(t_i)$$

then $P_N g$ is the orthogonal projection of g onto span $\{K_{t_i}\}_{i=1}^{N}$ and that $P_N g = g_N$, see Kimeldorf and Wahba [6]. In particular,

$$\min_a E(g(t)-\sum_{j=1}^{N} a_j g(t_j))^2 = \min_a ||K_t-\sum_{j=1}^{N} a_j K_{t_j}||_K \tag{1.1}$$

$$= \min_{a} \max_{\langle f,f\rangle_K \le 1} |f(t) - \sum_{j=1}^{N} a_j f(t_j)|.$$

Minimal norm interpolation also has the striking property that it furnishes the best estimator for g(t), $g \in H_K$ among *all* estimators (linear or nonlinear in $g(t_1),...g(t_N)$) which uses the information $g(t_1),...g(t_N)$ with $\langle g,g\rangle_K \le 1$, that is,

$$\min_{A} \max_{\langle f,f\rangle_K \le 1} |f(t) - A(f(t_1),....,f(t_N))|$$

where A is *any* map from $(f(t_1),...,f(t_N))$ into the real line is achieved for $A(g(t_1),...,g(t_N)) = g_N$. This property and various extensions and related matters in other normed spaces is described in C.A. Micchelli and T. J. Rivlin [13].

In each instance above the data is viewed is known exactly. Frequently in applications only "noisy" data is available and this leads to the problem of data smoothing. We briefly discuss this problem in Section 5.

As a criteria for choosing $t_1,...,t_N$ we minimize

$$(1.2) \qquad E\int_T (g(t) - g_N(t))^2 dt := J(T_N)$$

where the expectation is taken with respect to the prior covariance K(s,t). It is not hard to show that

$$(1.3) \qquad J = J(T_N) = \int_T \{K(t,t) - (K_{t_1}(t),...,K_{t_N}(t))K_N^{-1}(K_{t_1}(t),...K_{t_N}(t))'\}dt.$$

In practice K may have to be estimated by use of a finer trial grid of points than will ultimately be used, or from physical principles governing the phenomena under study. The covariance of air pollution measurements for example surely depends on the local geography. If K is known, then, frequently the minimization of J will have to be carried out numerically. In this paper we will provide a lower bound for J in terms of the eigenvalues associated with the integral operator induced by K. Thus, trial solutions for the design T_N minimizing J may

be compared against the lower bound to decide whether the further minimization is worth while.

Theorem 1. Let the operator K defined by $(Kf)(t) = \int_T K(t,s)f(s)ds$ *be a symmetric compact operator of* $L_2(T)$ *into itself and have eigenvalues* $\lambda_1 \geq \lambda_2 \geq \ldots$ *Then*

$$\inf_{T_N} J(T_N) \geq \sum_{i=N+1}^{\infty} \lambda_i = \int_T K(t,t)dt - \sum_{i=1}^{N} \lambda_i.$$

It is not known whether or not his lower bound can be achieved.

A fair amount is known about optimum designs for T= [0,1], see Sachs and Ylvisaker [15], Wahba [19], Hajek and Kimeldorf [3]. A sequential procedure for choosing an optimum design for T= [0,1] is given in Athavale and Wahba [1]. The sequential procedure depends heavily on properties of optimal designs known from the earlier papers and does not at present generalize to T= [0,1] x [0,1] or the sphere. In fact it appears that nothing is known about best possible convergence rates in several dimensions for $||g-P_N g||_K^2$, see Ylvisaker [21].

Sachs and Ylvisaker [16] have shown that $||g-P_N g||_k^{-2}$ is the variance of the Gauss-Markov estimate of θ in the model

$$Y(t) = \theta g(t) + X(t), \quad t \in T, \; EX(s)X(t) = K(s,t)$$

given $Y(t), t \in T_N$.

If it is known that g is in some class C then it might be desirable to choose T_N to minimize $\sup_{g \in C} ||g-P_N g||_K^2$. Through the notion of N-widths, introduced by Kolmogorov [7], and asymptotic estimates of the eigenvalues of certain integral operators we will provide lower bounds for the supremium of the design error for g in a certain class C. The class we will consider here is the natural generalization of the function class for which optimal one dimensional designs were obtained in [3, 15, 19]. Before stating this result we review briefly some results for optimal experimental design from [19] for $T = [0,1]$.

The basic assumption made in [19] about K is that it has the characteristic discontinuity of a Green's function for a 2m *th* order self-adjoint differential operator,

$$(1.3). \quad \alpha(t) = \lim_{s \downarrow t} \frac{\partial^{2m-1}}{\partial s^{2m-1}} K(s,t) - \lim_{s \uparrow t} \frac{\partial^{2m-1}}{\partial s^{2m-1}} K(s,t).$$

Suppose that g has a representation

$$g(t) = \int_0^1 K(t,s)\rho(s)ds, \qquad \rho \in L_2[0,1].$$

for some $\rho \in L_2[0,1]$ and let $T_N = \{t_{iN}\}_{i=1}^{N}$ be determined by a strictly positive density f,

$$\int_0^{t_{iN}} f(s)ds = i/(N+1,) \quad i = 1,2,\ldots,N; \quad N = 1,2,\ldots.$$

Under various regularity conditions including $\alpha, \rho > 0$

$$||g - P_{T_N} g||_K^2$$
$$= \frac{C_m}{N^{2m}} \left\{ \int_0^1 \frac{\rho^2(s)\alpha(s)}{f^{(2m)}(s)} ds \right\} (1 + o(1))$$

where C_m is a constant depending on m. The density f is chosen to minimize the quantity in brackets, see [19]. Thus the rate of decay of $||g - P_{T_N} g||_K^2$ is asymptotically the same as the decay of the eigenvalues of K (see Naimark [14]).

We return now to a general set T and reproducing kernel K.

Theorem 2. Let K be a symmetric compact operator from $L_2(T)$ into itself with eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots$. Let $C = \{g : g(t) = \int_T K(t,s)\rho(s)ds, \int_T \rho^2(s)ds \leq 1\}$.

Then

$$\sup_{g \in C} ||g - P_N g||_K^2 \geq \lambda_{N+1}$$

Next, (Section 3) we investigate the eigensequences for certain useful reproducing kernels on $T = \otimes^d[0,1]$, the tensor product of [0,1] d times. We will prove

Theorem 3. Let $H_K = \otimes^d H_Q$, *where* H_Q *is an r.k.h.s. on* $[0,1]$ *with* Q *satisfying (1.3). Then*

$$(1.3) \qquad \lambda_N = 0((\log N)^{d-1}/N)^{2m}.$$

Based on this result, we make some conjectures (Section 4) concerning good designs in H_K using results from the multi-dimensional quadrature literature. In particular, we conjecture (1.3) is the optimal rate, which has only been proved for d=1, as explained above. Finally in section 5 we make some observations concerning noisy data.

2. Lower bounds for optimal designs.

We begin with the proof of

Theorem 1. Let the (symmetric) operator K defined by $(Kf)(t) = \int_T K(t,s)f(s)ds$ *be compact with eigenvalues* $\lambda_1 \geq \lambda_2 \geq \ldots$. *Then*

$$\inf_{T_N} J(T_N) = \inf_{T_N} \int_T \| K_t - P_N K_t \|_K^2 dt \geq \sum_{\nu=N+1}^{\infty} \lambda_\nu .$$

Proof: **The equality is immediate from (12). Since**

$$\int_T \| K_t - P_N K_t \|_K^2 dt$$

$$= \int_T \| K_t \|^2 dt - \int_T \| P_N K_t \|_K^2 dt$$

$$= \sum_{\nu=1}^{\infty} \lambda_\nu - \int_T \| P_N K_t \|_K^2 dt,$$

it suffices to show that $\int_T \| P_N K_t \|_K^2 dt \leq \sum_{\nu=1}^{N} \lambda_\nu$.

Let $\phi_1, \ldots, \phi_N$ be any N orthonormal functions in H_K. Then the projection of K_t onto span $\{\phi_i\}_{i=1}^N$ is

$$P_N K_t = \sum_{i=1}^{N} \phi_i(t)\phi_i,$$

and

$$\int_T ||P_N K_t||_K^2 = \sum_{i=1}^{N} ||\phi_i||_{L_2}^2.$$

Let $K^{1/2}$ be the symmetric square root of the operator K. Then by the properties of the reproducing kernel norm,

$$||\phi_i||_{L_2}^2 = ||K^{1/2}\phi_i||_K^2$$

Now by the extremal properties of the eigenvalues of K,

$$\sup_{\phi \in H_K} ||K^{1/2}\phi||_K^2 / ||\phi||_K^2 = \lambda_1,$$

$$\sup_{\substack{\phi \in H_K \\ (\phi,\psi_1)_{L_2}=0}} ||K^{1/2}\phi||_K^2 / ||\phi||_K^2 = \lambda_2,$$

where ψ_1 is the maximizing element for the first equality above, etc. Thus,

$$\sum_{i=1}^{N} ||\phi_i||_{L_2}^2 \leq \sum_{\nu=1}^{N} \lambda_\nu.$$

This result is also a consequence of a classical result from the theory of integral equations (see [18], 149).

Theorem 2. Let H_N be any N dimensional subspace in H_K and P_N be the orthogonal projection onto H_N. Then there exists a function g,

$$g(t) = \int_T K(t,s)\rho(s)ds,$$

such that

$$||g - P_{H_N} g||_K^2 \geq \lambda_{N+1} \int_T \rho^2(s)ds \tag{2.1}$$

Proof: The proof of this theorem also follows directly from the external properties for the eigenvalues of $K^{1/2}$ and has an interpretation in the theory of N-widths, [17]. Specifically we have

$$\inf_{H_N} \sup_{g \in C} ||g - P_{H_N} g||_K^2$$

$$= \inf_{H_N} \sup_{||\rho||_{L_2}=1} ||K\rho - P_{H_N} K\rho||_K^2$$

$$= \inf_{H_N} \sup_{||\rho||_{L_2}=1} ||K^{1/2}\rho - P_{H_N} K^{1/2}\rho||_{L_2}^2.$$

The extremal properties of eigenvalues and eigenfunctions of symmetric operators imply that

$$\sup_{||\rho||_{L_2}=1} ||K^{1/2}\rho - P_{H_N} K^{1/2}\rho||_{L_2}^2.$$

achieves its minimum for H_N equal to the span of the first N eigenfunctions of $K^{1/2}$ and the value of the minimum is λ_{N+1}.

To prove the existence of an optimal design we must find a subspace of the form span $\{K_t : t \in T_N\}$ for some design T_N which achieves the lower bound in (2.1) which would be expected to be close to the span of the first n eigenfunctions.

It should not be expected that for an arbitrary covariance kernel K an optimal design exists for each N. However, for certain classes of kernels existence of optimal designs has been shown, see Melkman [10]; Melkman and Micchelli [11].

3. Good designs in Tensor Product Spaces

To make use of Theorem 2 we will obtain the asymptotic rate of decay of the eigenvalues of the r.k. for H_K of the form

$H_K = \otimes^d H_Q$ (tensor product of d copies of H_Q)

where H_Q is an r.k.h. s of functions on $[0,1]$ with eigenvalues that decay as a power, $\lambda_\nu = c\nu^{-2m}(1 + o(1)), \nu \to \infty$. For instance, if Q behaves as a Green's function for a 2mth order linear differential operator this condition is satisfied. As a simple example of this possibility, let $H_Q = \{f: f, f', \ldots, f^{(m-1)}$ abs. cont. $f^{(m)} \in L_2[0,1], f^{\nu}(0) = f^{\nu}(1), \nu = 0,1 \ldots, m-1\}$ with inner product,

$$<f,g> = \left(\int_0^1 f(u)du\right)\left(\int_0^1 g(u)du\right) + \int_0^1 f^{(m)}(u)g^{(m)}(u)du.$$

Then the r. k. Q is

$$Q(s,t) = 1 + \sum_{\nu \neq 0} \frac{e^{2\pi i\nu(s-t)}}{(2\pi\nu)^{2m}}$$

and the corresponding eigenvalues and eigenfunctions are

$$\{1, (2\pi\nu)^{-2m} : \nu = \pm 1, \ldots\}, \quad \{e^{2\pi i\nu s};\ \nu = 0, \pm 1, \}.$$

Theorem 3. Let $H_K = \oplus^d H_Q$ *where the eigenvalues* $\{\lambda_\nu\}$ *of* H_Q *satisfy* $\lambda_\nu = \nu^{-2m}(1 + o(1))$ *then the eigenvalues* $\{\xi_\nu\}$ *of* H_K *satisfy*

$$\xi_N = \left(\frac{(\log N)^{d-1}}{N}\right)^{2m}(1 + o(1))$$

Proof. Since $K = \oplus^d Q$, the eigenvalues of K are the tensor product of the eigenvalues of Q i.e. if $\xi_1 \geq \xi_2 \geq \ldots$ are the eigenvalues of K then

$$\{\xi_N\} = \{\lambda_{j_1}\lambda_{j_2} \cdots \lambda_{j_d} :\ (j_1, \cdots j_d),\ j_1, \cdots j_d = 1,2,\ldots\}$$

To estimate the decay of ξ_N we observe that the number of lattice points $(j_1, \ldots j_d)$ satisfying $\prod_{\ell=1}^{d} j_\ell \leq k$ is $k(\log k)^{d-1}(1 + o(1))$.

Hence, since $\lambda_\nu = \nu^{-2m}(1 + o(1))$ we have

$$\xi_{[k(\log k)^{d-1}]} = k^{-2m}(1 + o(1)).$$

($[x]$ = greatest integer $\leq x$). Choosing $k = [N(\log N)^{d-1}]$ gives the desired conclusion,

$$\xi_N = \left(\frac{(\log N)^{d-1}}{N}\right)^{2m} (1 + o(1)).$$

It is not known for $d>1$ whether there exists a design for which

$$\| g - P_N g \|_K^2 \leq \text{const}\left(\frac{(\log N)^{d-1}}{N}\right)^{2m} \int_{T_N} \rho^2(u)du.$$

However designs with a convergence rate approaching the optimum rate have been given in Wahba [20] for $d=2$.

Define

$$Z_n^j = \left\{\frac{k}{n^j} : k = 1,\ldots,n^j\right\}$$

and

$$T_{N,\ell} = \bigcup_{j=1}^{\ell+1} Z_n^j \otimes Z_n^{\ell+1-j}.$$

In [20] it is shown that $T_{N,\ell}$ has $N = (\ell + 1)n^{\ell+2} - \ell n^{\ell+1}$ distinct points and for $H_K = H_Q \otimes H_Q$.

$$(3.1) \qquad \| g - P_{T_{N,\ell}} g \|^2 \leq \text{const}\,\frac{(\ell + 2)^2}{n^{(\ell+1)2m}}\left(\int_T \rho^2(u)du\right)(1 + o(1))$$

$$\leq \text{const}\,\frac{(\ell + 2)^{2+\left(\frac{\ell+1}{\ell+2}\right)2m}}{N^{\left(\frac{\ell+1}{\ell+2}\right)2m}}\left(\int_T \rho^2(u)du\right)(1 + 0(1))$$

where $g = K\rho$ and $o(1) \to 0$ as $n \to \infty$

Choosing $\ell = (\log N)^p(1 + o(1))$ for any p, $0 < p < 1$ we have

$$N = (\ell + 1)n^{\ell+2}(1 + o(1))$$

or $\log N = p \log \log N + (\log N)^p \log n(1 + o(1))$. Hence

$$\log n = (1 + o(1))\frac{\log N - p \log \log N}{(\log N)^p}$$

and n $\to\infty$ provided $0<p<1$ (for $p=1$ this conclusion fails). Setting $p = \frac{m}{m+1}$ into (3.1) gives

$$||g-P_N g||_K^2 = 0\left(\frac{(\log N)^{\left(1+\frac{\ell+1}{\ell+2}m\right)/(m+1)}}{N^{\frac{\ell+1}{\ell+2}}}\right)^{2m},$$

a convergence rate which approaches the optimum rate of $(\log N/N)^{2m}$ implied by Theorems 2 and 3.

4. Optimal quadrature - a conjecture

A quadrature formula for $\int_T g(t)dt$ can be obtained by setting $\int_T (P_N g)(t)dt = \sum_{i=1}^{N} c_i g(t_i)$. Then

$$\begin{aligned}
&|\int_T g(t)dt - \int_T (P_N g)(t)dt| \\
&= |\langle \eta, g - P_N g\rangle_K| \\
&= |\langle \eta - P_\eta \eta, g - P_N g\rangle_K| \\
&\le ||\eta - P_N\eta||_K\, ||g-P_N g||_K \\
&\le ||\eta - P_N\eta||_K\, ||g||_K
\end{aligned}$$

where η is the representer of integration in H_K.

$$\eta(s) = \langle \eta, K_s\rangle_K = \int_T K(s,u)du$$

An optimal quadrature problem may be formulated as: Find $t_1,\ldots,t_N$ to minimize $||\eta - P_N\eta||_K$.

There is a large literature on choosing sequences in the d-dimensional unit cube which makes the error for the special quadrature formula $\frac{1}{N}\sum_{i=1}^{N} g(t_i)$ asymptotically small. This work has focused on finding sequences $T_N = \{t_1,\ldots,t_N\}$ for which the discrepancy D_N defined by

$$D_N = \sup_t |F_N(t) - F(t)|$$

is small. Here F_N is the cumulative distribution function of the point set and F is the cumulative distribution of the uniform density, see Kuipers and Neiderreiter [8], Halton [4], Halton and Zaremba [5], Zaremba [22].

It is known that the Hammersly sequences defined below, have discrepancy

$$D_N = \left(\frac{\log^{d-1} N}{N}\right)(1 + o(1))$$

see Halton [4].

These sequences are defined (in d-dimensions) by

$$\left\{\frac{(n}{N}, \phi_2(n), \phi_3(n), \ldots, \phi_{p_d}(n))\right\}_{n=0}^{N-1}$$

where the subscripts in the ϕ's are successive primes and if $n = \sum_{j=o}^{M} n_j p^j$ where $M = [\log_p n]$,

then $\phi_p(n) = \sum_{j=1}^{M+1} n_j p^{-j}$.

Bounds on $\epsilon_N = |\int_T g(t)dt - \frac{1}{N} \sum_{i=1}^{N} g(t_i)|$

in terms of the discrepancy appear in the literature, see Kuipers and Neiderreiter [8, p. 157], Zaremba [22] and references therein.

In [8] it is shown for certain sequences that $\epsilon_N = 0(D_N^q)$ where g has

Fourrier coefficients $c_{\nu_1, \nu_2, \ldots, \nu_d}$ satsifying

$$|c_{\nu_1, \nu_2, \ldots, \nu_d}| \leq \frac{M}{\left(\prod_{i=1}^{d} \bar{\nu}_i\right)^q}$$

where

$$\bar{\nu}_i = \begin{cases} |\nu_i| & \nu_i \neq 0 \\ 1 & \nu_i = 0 \end{cases}.$$

We conjecture that similar results obtain for Hammersley sequences and that for spaces H_K satisfying the hypothesis of Theorem 2 the optimal convergence rate

$$||g - P_N g||_K^2 = \lambda_{N+1}(1 + o(1))$$

$$= \left(\frac{(\log N)^{d-1}}{N}\right)^{2m}(1 + o(1))$$

will hold for T_N the Hammersly sequence and g of the form $g=K\rho$.

5. Noisy Data

In this section we include some remarks concerning estimation based on inaccurate data. If instead of observing $g(t), t\in T_N$, we assume that the data is given by

$$y_i = g(t_i) + \epsilon_i, \qquad E\epsilon_i\epsilon_j = \delta_{ij}.$$

then the minimum norm estimator

$$\min_a E(g(t) - \sum_{i=1}^N a_i y_i)^2 = ||K_t - \sum_{j=1}^N a_j K_{t_j}||_K^2 + \sigma^2 \sum_{j=1}^N a_j^2$$

leads in the functional-analytical approach to

$$(5.1) \qquad = \min_a \max_{<f,f>_K \le 1} |f(t) - \sum_{i=1}^N a_i f(t_i)|^2 + \sigma^2 \sum_{i=1}^N a_j^2$$

Recently, this variational problem has been solved by Laurent [9]. It has been shown that the minimum norm estimator is the smoothing "spline" in H_K with parameter σ^{-2}, that is, if g_N minimizes

$$(5.2) \qquad \min \{ ||f||_K^2 + \sigma^{-2} \sum_{i=1}^N (f(t_i) - g(t_i))^2 \}$$

then $g_N(t) = \sum_{i=1}^N c_i(t) g(t_i)$ is the minimum norm estimator for g when we have noisy data. Note that the smoothing parameter σ^2 does *not* depend on the value $t \in T$ at which we choose to estimate g(t). The following short proof of this result is instructive: We wish to determine

$$\min_a \| K_t - \sum_{i=1}^{N} a_i K_{t_i} \|_K^2 + \sigma^2 \sum_{j=1}^{N} a_j^2$$

To this end, we introduce the tensor product space $H_K \otimes R^N = \{(f,a) \mid f \in H_K, a \in R^N\}$ with the norm

$$\| (g,a) \|_\sigma^2 = \| g \|_K^2 + \sigma^2 \sum_{i=1}^{N} a_i^2$$

Then the above problem in $H_K \otimes R^N$ becomes

$$\min_a \| h - \sum_{j=1}^{N} a_j h_j \|_\sigma^2$$

$h = (K_{t},0)$, $h_j = (K_{t_j}, -e_j)$ $\quad (e_j)_k = \delta_{jk}$.

But from the theory for estimating exactly given data, as in (1.1), the minimum $a = (a_1, \ldots, a_N)$ may be obtained from the best interpolant,

$$\min_{\langle (f,a),(K_{t_i},-e_i)\rangle = g(t_i)} \| (f,a) \|_\sigma^2 =$$

$$\min_f \{ \| f \|_K^2 + \sigma^{-2} \sum_{i=1}^{N} (g(t_i) - f(t_i))^2 \}$$

in agreement with (5.2).

It has not yet been determined if the optimality of smoothing "splines" persists when an estimator for the full function g(t), $t \in T$ when the error criteria (1.2) is used. However, let us replace (5.1) by

$$\min_a \max_{\substack{\langle f,f\rangle \le 1 \\ \sum_{i=1}^{N} r_i^2 \le 1}} | f(t) - \sum_{i=1}^{N} a_i (f(t_i) + \epsilon r_i) |^2$$

that is, we minimize the worst least square error when we know the noise in the data is in the region $y_i = f(t_i) + \epsilon r_i$, $\sum_{i=1}^{N} r_i^2 \le 1$. It has been shown that in this setting the smoothing spline is also optimal. However, unlike (5.1), the smoothing parameter *depends* on ϵ as well as $t \in T$. Moreover, this theory holds in great generality, including, in particular, estimating the full function g(t), $t \in T$ (see Melkman and Micchelli [13] for the details). For methods of

choosing the smoothing parameter using a cross-validation procedure based on the data see Craven and Wahba [2].

The design problem of Theorem 1, has an analogue for noisy data which may be described as follows. Let g(t), t ϵ T be a stochastic process as before and let $g_N(t) =$

$$E\{g(t)\,|\,y_1,\ldots,y_N\} = (K_{t_1}(t),\ldots,K_{t_N}(t))(K_N + \sigma^2 I)^{-1}(y_1,\ldots y_N)'.$$

Then J_N becomes

$$J_N = E\int_T (g(t)-g_N(t))^2 dt$$

$$= \int_T \{K(t,t)-(K_{t_1}(t)\ldots,K_{T_N}(t))(K_N + \sigma^2 I)^{-1}(K_{t_1}(t),\ldots,K_{t_N}(t))'dt.$$

which may be compared to equation (1.2).

References

1. Athavale, M., and Wahba, G. (1978), Determination of an optimal mesh for a collocation-projection method for solving two-point boundary value problems. Department of Statistics, University of Wisconsin, Madison T. R. No. 530, to appear *J. Approx. Theory.*

2. Craven, P., and Wahba, G., (1979) Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation, *Numer. Math.* 31, 377-403.

3. Hajek, J. and Kimeldorf, G. (1974). Regression designs in autoregressive stochastic processes. *Ann. Statist 2, 520-527.*

4. Halton, J. H., (1960). On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals, *Numer. Math.* 2, 84-90.

5. Halton, J. H. and Zaremba, S. K. (1969), The extreme and L^2 discrepancies of some plane sets. *Monatshefte für Mathematik 73, 316-328.*

6. Kimeldorf, G. S. and Wahba, G. (1970) A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. *Ann. Math. Statist.*, 41,2.

7. Kolmogorov, A., (1936) über die best annäherung von funktioner einer gegebenen funktionenklasse, *Ann. of Math,* vol. 37, pp. 107-110.

8. Kuipers, L. and Neiderreitter, H. (1974) "Uniform distribution of sequences," John Wiley & Sons, New York.

9. Laurent, P. J., Colloquium talk, Univ. of Wisconsin, Department of Statistics, April 1978.

10. Melkman A. A. (1976) n-widths and optimal interpolation of time-and band-limited functions, in "Optimal Estimation in Approximation Theory," C. A. Michelli and T. J. Rivlin, eds, Plenum Press, New York, 1977.

11. Melkman, A. A. and C. A. Micchelli, (1978) Spline spaces are optimal for L^2 n-widths, *Illinois Journal of Mathematics,* 22, 541-564.

12. Melkman, A. A. and C. A.Micchelli, Optimal estimation of linear operators in Hilbert spaces from inaccurate data, to appear *SIAM Journal of Numer. Anal.*

13. Micchelli, C. A. and T. J. Rivlin, A survey of optimal recovery, in"Optimal Estimation in Approximation Theory," C. A. Micchelli and T. J. Rivlin, Eds., Plenum Press, New York 1977.

14. Naimark, M.A. (1967) Linear Differential Operators, Frederick Ungar Publishing Co., New York.

15. Sachs, J. and Ylvisaker, D. (1969), Designs for regression problems with correlated errors. *Ann. Math. Statist.* 49, 2057-2074.

16. Sachs, J. and Ylvisaker, D., (1970), Statistical designs and integral approximation, Proc. of the 12th Biennial Seminar of the Canadian Mathematical Congress.

17. Shapiro, Harold, (1971) Topics in Approximation, Lecture Notes in Mathematics, 187, Springer, New York, p. 187.

18 Smithies, F. (1965) "Integral Equations," Cambridge Tracts in Mathematics and Mathematical Physics, No. 49, Cambridge University Press.

19. Wahba, G. (1971), On the regression design problem of Sachs and Ylvisaker, *Ann. Math. Statist.* 42, 3, 1035-1053.

20. Wahba, G. (1978), Interpolating surfaces: High order convergence rates and their associated designs,with application to X-ray image reconstruction. Department of Statistics, University of Wisconsin - Madison T.R. No. 523, to appear in *SIAM J. Numer.*

21. Ylvisaker, D. (1975), Designs on random fields, in "A survey of Statistical Design and Linear Models, J. Sivasrava, ed., North-Holland 593-607.

22. Zaremba, S. K. (1968). The Mathematical basis of Monte Carlo and quasi-Monte Carlo methods. *SIAM Review 3, 303-314.*

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Technical Report No. 565	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) 6 DESIGN PROBLEMS FOR OPTIMAL SURFACE INTERPOLATION		5. TYPE OF REPORT & PERIOD COVERED
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) 10 Charles A. Micchelli ~~and~~ Grace Wahba		8. CONTRACT OR GRANT NUMBER(s) 15 N00014-77-C-0675
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics University of Wisconsin 1210 West Dayton Street, Madison, WI 53706		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 12 6p.
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research 800 N. Quincy Street Arlington, VA 22217		12. REPORT DATE 11 May 1979
		13. NUMBER OF PAGES 19
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) 14 UWIS-DS-79-565		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Distribution of this document is unlimited 9 Technical rept.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) experimental design, surface interpolation, optimal design, reproducing kernel Hilbert spaces, optimal approximation		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) We consider the problem of interpolating a surface given its values at a finite number of points. We place a special emphasis on the question of choosing the location of the points where the function will be sampled. (continued on next page)		

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Using minimal norm interpolation in reproducing kernel Hilbert spaces, equivalently Bayesian interpolation, and N-widths, we provide lower bounds for interpolation error relative to certain error criteria. These lower bounds can be used when evaluating an existing design, or when attempting to obtain a good design by iterative procedures to decide whether further minimization is worthwhile. The bounds are given in terms of the eigenvalues of a relevant reproducing kernel and the asymptotic behavior of these eigenvalues for certain tensor product spaces in the unit d-dimensional cube is obtained.

We demonstrate that for H_m, the d-dimensional tensor product of Sobolev spaces $W_2^{(m)}[0,1]$ and $P_N g$, the minimal norm interpolant to g at N given data points, the uniform convergence of $\| g - P_N g \|_{H_m}$ over g in the unit ball in H_{2m} cannot proceed at a rate faster than $((\log N)^{d-1}/N)^{2m}$. Certain conjectures concerning designs converging at this rate are made.

END

DATE
FILMED
1-82

DTIC